

WHAT IS CLAIMED IS:

- 1 1. A method of creating a generic text summary of a document; said method  
2 comprising:
  - 3 creating a weighted document term-frequency vector for said  
4 document;
  - 5 for each sentence in said document, creating a weighted sentence  
6 term-frequency vector;
  - 7 computing a score for each said weighted sentence term-frequency  
8 vector in accordance with relevance to said weighted document  
9 term-frequency vector; and
  - 10 selecting a sentence for inclusion in said generic text summary in  
11 accordance with said computing.
- 1 2. The method of claim 1 further comprising:
  - 2 deleting said sentence from said document and eliminating terms in  
3 said sentence from said document;
  - 4 recreating said weighted document term-frequency vector in  
5 accordance with said deleting and said eliminating; and
  - 6 selectively repeating said computing, said selecting, said deleting,  
7 said eliminating, and said recreating.
- 1 3. The method of claim 2 wherein said selectively repeating is terminated  
2 when a predetermined number of sentences has been selected.
- 1 4. The method of claim 1 wherein said computing comprises calculating an  
2 inner product of said weighted sentence term-frequency vector and said weighted  
3 document term-frequency vector.
- 1 5. The method of claim 1 wherein said creating a weighted sentence term-  
2 frequency vector comprises implementing a local weighting function and  
3 implementing a global weighting function.
- 1 6. The method of claim 5 wherein said creating a weighted sentence term-  
2 frequency vector comprises normalizing each said weighted sentence term-  
3 frequency vector.

1 7. The method of claim 1 wherein said creating a weighted document term-  
2 frequency vector comprises implementing a local weighting function and  
3 implementing a global weighting function.

1 8. The method of claim 7 wherein said creating a weighted document term-  
2 frequency vector comprises normalizing said weighted document term-frequency  
3 vector.

1 9. A system for creating a generic text summary of a document; said system  
2 comprising:

3 a computer;  
4 a display for displaying said generic text summary; and  
5 summarizer program code, operable on said computer, for  
6 analyzing and summarizing said document; said summarizer  
7 program code comprising:  
8 a vector generator for creating a weighted document term-  
9 frequency vector for said document and creating a weighted  
10 sentence term-frequency vector for each sentence in said  
11 document;  
12 a scoring engine for computing a score for each said weighted  
13 sentence term-frequency vector in accordance with relevance to  
14 said weighted document term-frequency vector; and  
15 a selector for selecting a sentence for inclusion in said generic text  
16 summary in accordance with output results from said scoring  
17 engine.

1 10. The system of claim 9 wherein said summarizer program code further  
2 comprises a document editor for deleting said sentence from said document and  
3 for eliminating terms in said sentence from said document; and wherein said  
4 vector generator recreates said weighted document term-frequency vector in  
5 accordance with output results from said document editor.

1 11. The system of claim 10 wherein said summarizer further comprises a loop  
2 routine for generating iterative sequential operations of said vector generator, said  
3 scoring engine, said selector, and said document editor.

1 12. The system of claim 11 wherein said loop routine is responsive to a  
2 predetermined limit such that said generic text summary is of a predetermined  
3 number of sentences.

1 13. A method of creating a generic text summary of a document; said method  
2 comprising:

3 decomposing said document into individual sentences;  
4 forming a candidate sentence set from said individual sentences;  
5 for each of said individual sentences in said candidate sentence set,  
6 creating a weighted sentence term-frequency vector;  
7 creating a weighted document term-frequency vector for said  
8 document;  
9 for each of said individual sentences in said candidate sentence set,  
10 computing a relevance score for said weighted sentence term-  
11 frequency vector relative to said weighted document term-  
12 frequency vector;  
13 selecting a sentence for inclusion in said generic text summary in  
14 accordance with said computing;  
15 deleting said sentence from said candidate sentence set;  
16 eliminating terms in said sentence from said document; and  
17 recreating said weighted document term-frequency vector in  
18 accordance with said deleting and said eliminating.

1 14. The method of claim 13 further comprising:  
2 selectively repeating said computing, said selecting, said deleting,  
3 said eliminating, and said recreating.

1 15. The method of claim 14 wherein said selectively repeating is terminated  
2 when a predetermined number of sentences has been selected.

1 16. The method of claim 13 wherein said computing comprises calculating an  
2 inner product of said weighted sentence term-frequency vector and said weighted  
3 document term-frequency vector.

1 17. The method of claim 13 wherein said creating a weighted sentence term-  
2 frequency vector comprises implementing a local weighting function and  
3 implementing a global weighting function.

1 18. The method of claim 17 wherein said creating a weighted sentence term-  
2 frequency vector comprises normalizing each said weighted sentence term-  
3 frequency vector.

1 19. The method of claim 13 wherein said creating a weighted document term-  
2 frequency vector comprises implementing a local weighting function and  
3 implementing a global weighting function.

1 20. The method of claim 19 wherein said creating a weighted document term-  
2 frequency vector comprises normalizing said weighted document term-frequency  
3 vector.

1 21. A method of creating a generic text summary of a document; said method  
2 comprising:  
3 constructing a terms-by-sentences matrix for said document;  
4 performing singular value decomposition on said terms-by-  
5 sentences matrix to obtain a singular value matrix and a right  
6 singular vector matrix, wherein each sentence in said document is  
7 represented by a column vector of a transpose of said right singular  
8 vector matrix;  
9 ranking each right singular vector in said right singular vector  
10 matrix; and

11 selecting a sentence for inclusion in said generic text summary in  
12 accordance with said ranking.

1 22. The method of claim 21 further comprising repeating said selecting.

1 23. The method of claim 22 wherein said repeating is terminated when a  
2 predetermined number of sentences has been selected.

1 24. The method of claim 21 wherein said selecting further comprises  
2 identifying a sentence having a desired index value with said right singular vector.

1 25. The method of claim 21 wherein said constructing comprises  
2 implementing a local weighting function and implementing a global weighting  
3 function.

1 26. A system for creating a generic text summary of a document; said system  
2 comprising:

3 a computer;  
4 a display for displaying said generic text summary; and  
5 summarizer program code, operable on said computer, for  
6 analyzing and summarizing said document; said summarizer  
7 program code comprising:  
8 a matrix generator for creating a terms-by-sentences matrix for said  
9 document;  
10 an SVD performer for performing singular value decomposition on  
11 said terms-by-sentences matrix to generate a singular value matrix  
12 and a right singular vector matrix;  
13 a vector analyzer for ranking each sentence in said terms-by-  
14 sentences matrix in accordance with index values with said right  
15 singular vector matrix; and  
16 a selector for selecting a sentence for inclusion in said generic text  
17 summary in accordance with output results from said vector  
18 analyzer.

1 27. The system of claim 26 wherein said summarizer program further  
2 comprises a loop routine for generating iterative operation of said selector.

1 28. The system of claim 27 wherein said loop routine is responsive to a  
2 predetermined limit such that said generic text summary is of a predetermined  
3 number of sentences.

1 29. A method of creating a generic text summary of a document; said method  
2 comprising:

3 decomposing said document into individual sentences;  
4 forming a candidate sentence set from said individual sentences;  
5 constructing a terms-by-sentences matrix for said document;  
6 performing singular value decomposition on said terms-by-  
7 sentences matrix to obtain a singular value matrix and a right  
8 singular vector matrix, wherein each sentence in said candidate  
9 sentence set is represented by a column vector of a transpose of  
10 said right singular vector matrix;  
11 identifying a right singular vector from said right singular vector  
12 matrix;  
13 selecting a sentence for inclusion in said generic text summary in  
14 accordance with said identifying; and  
15 selectively repeating said identifying and said selecting.

1 30. The method of claim 29 wherein said selectively repeating is terminated  
2 when a predetermined number of sentences has been selected.

1 31. The method of claim 29 wherein said selecting further includes identifying  
2 a sentence in said candidate sentence set having a desired index value with said  
3 right singular vector.

1 32. The method of claim 29 wherein said constructing comprises  
2 implementing a local weighting function and implementing a global weighting  
3 function.